Heterogeneity in Rates of Recombination Across the Mouse Genome

by

Michael W. Nachman
and
Gary A. Churchill

BU-1305-M

September 1995

Heterogeneity in rates of recombination across the mouse genome

Michael W. Nachman* and Gary A. Churchill†

*Section of Genetics and Development

Biotechnology Building

Cornell University

Ithaca, NY 14853

†Biometrics Unit

Warren Hall

Cornell University

Ithaca, NY 14853

Running head: Recombination in mice

Send correspondence to:

Michael W. Nachman

Section of Genetics and Development

Biotechnology Building

Cornell University

Ithaca, NY 14853

Phone: (607) 254-4839

Fax: (607) 255-6249

E-mail: MWN1@cornell.edu

Abstract

Lyon (1976) suggested that the distribution of markers on a linkage map may be informative for identifying variation in regional rates of recombination throughout a genome even if the physical location of markers is unknown. If genes are randomly located physically, then regions of low recombination should contain clusters of loci on a genetic map. We have used this idea to estimate recombination rates from the *Drosophila melanogaster* linkage map. These results were compared with the results of two other studies which estimated regional recombination rates in *D. melanogaster* using both physical and genetic maps (Kliman and Hey 1993; Kindahl and Aquadro 1995). The three methods were largely concordant in identifying large-scale genomic patterns of recombination, particularly in regions of low recombination. The concordance between Lyon's method and each of the other two methods was as good as the concordance between the other two methods, suggesting that Lyon's method provides a reasonable approximation to recombinational patterns. Lyon's method was then applied to the *Mus musculus* microsatellite linkage map. The distribution of markers on this map was non-random and provided evidence for heterogeneity in recombination rates. Centromeric regions for several mouse chromosomes had significantly greater numbers of markers than expected under a homogeneous distribution, suggesting that recombination rates were lower in these regions. In contrast, most telomeric regions contained significantly fewer markers than expected. This suggests that recombination rates are elevated at the telomeres of most mouse chromosomes, and is corroborated by a comparison of the mouse genetic and cytogenetic maps. Lyon's (1976) method may be a generally useful way to estimate variation in recombination rates across a genome in species for which genetic, but not physical, maps are available.

Introduction

Recombination plays a central role in genetics, yet little is known of the rate, pattern, or regulation of recombination in higher eukaryotes. On a fine scale, data from both mice and humans indicate that in some genes recombination is not random, but instead occurs in hotspots (e.g. Lebo et al. 1983; Steinmetz et al 1986; Grimm et al. 1989; Bryda et al 1992). On a genomic scale, comparison of the cytogenetic and genetic maps of *Drosophila melanogaster* reveals widespread variation in rates of recombination among different chromosomal regions (Lindsley and Sandler 1977; Kliman and Hey 1993; Kindahl and Aquadro, 1995). Genome-wide variation is observable in *D. melanogaster* because detailed cytogenetic and genetic maps exist for the same set of markers. Variation in recombination rate at the level of the genome has not been investigated as extensively in other species.

Lyon (1976) pointed out that it should be possible to infer variation in recombination rates along chromosomes from a genetic map alone if certain conditions are met. In particular, if loci are randomly distributed physically and if chiasmata are also randomly located, then loci should be randomly distributed on a linkage map. However, if chiasmata are not randomly distributed, but occur preferentially in certain regions, then this should be reflected in the distribution of loci on a linkage map. Regions with low levels of recombination should contain clusters of markers, while regions of high recombination should contain markers that are more widely spaced.

Using this approach, Lyon (1976) documented that the 268 markers then assigned to the laboratory mouse genetic map (Green 1975) were not randomly distributed. Instead, they occurred in clusters, with regions containing few markers between clusters, an observation interpreted as evidence for variation in recombination rates. Although Lyon's method is attractive because of its simplicity and potential applicability to any species for which a reasonably detailed genetic map exists, it has never been tested.

Here, Lyon's method is tested with the *D. melanogaster* linkage data. Recombination rates are estimated directly from the density of markers across the genetic map. These estimates of recombination rate are then compared with two different recently published estimates of regional recombination rates in *D. melanogaster*, both of which rely on comparisons of the physical (cytogenetic) and genetic maps (Kliman and Hey 1993; Kindahl and Aquadro, 1995).

The rapid increase in linkage data for the laboratory mouse, *Mus musculus*, also makes it possible to greatly extend the approach used by Lyon. The mouse genome is comprised of 19 acrocentric pairs of autosomes and acrocentric X and Y chromosomes. There are several distinct genetic maps for the mouse, maintained in separate databases and generated in different ways. One (referred to here as the gene-based map) is available on-line from the Mouse Genome Database (MGD). This map is a composite generated from different crosses in different laboratories, integrated with a set of anchor loci. The map includes phenotypic mutants, allozyme variants, cloned genes, and a variety of polymorphic molecular markers. Initially, many of the markers were placed using two- and three-point crosses between laboratory strains or recombinant inbred strains. Since the mid 1980's, however, many markers have been placed using interspecific backcrosses involving *Mus spretus* (Avner et al. 1988; Copeland and Jenkins 1991). A large framework of loci for this gene-based map has been generated at Frederick, Maryland from crosses of (C57BL/6J X *Mus spretus*)F_1 X C57BL/6J mice (Copeland et al. 1993). In this cross, female F_1's are backcrossed to male C57BL/6J *M. musculus* mice since male F_1's are sterile. Thus, only female meioses contribute to recombinational distances. Markers situated from these crosses are placed on the composite map using anchor loci.

Another mouse genetic map (Dietrich et al. 1992; Dietrich et al. 1994) was derived from simple sequence length polymorphisms (SSLP) or microsatellites (referred to here as the microsatellite map), and is available on-line as the "Whitehead Institute/MIT Genome Center genetic map of the mouse." Random clones containing the repeat sequence $(CA)_n$

were screened by oligonucleotide hybridization to a small-insert mouse genomic library. A small number of additional microsatellites were identified from known gene sequences. PCR primers for microsatellites were designed to lie in single-copy sequence surrounding the variable, dinucleotide repeats. The map was constructed using a single (OB x CAST) F_2 intercross and thus reflects sex-averaged genetic distances from two laboratory strains of *Mus musculus*. The current map contains approximately 6,000 markers and has a total length of approximately 1,400 cM. The gene-based map and microsatellite map have been partially integrated using a subset of the microsatellites in the Frederick interspecific backcross (Copeland et al. 1993).

Differences between the gene-based map and the microsatellite map are expected for at least three reasons (Copeland et al. 1993; Dietrich et al. 1992, 1994). First, recombinational distances are often different in male and female meioses, with female distances typically, though not always, being greater (Davisson et al. 1989). Distances estimated from female backcrosses to *M. spretus* may consequently be different from sex-averaged distances measured in intercrosses. Second, it is possible that recombination rates in *M. spretus* may be different from those in *M. musculus* for some genomic regions, although comparisons of markers mapped using both approaches reveal generally good agreement (Avner et al. 1988). Third, the microsatellite map consists mostly of markers chosen at random while the gene-based map includes markers chosen because of interest in a specific gene or chromosomal region. Thus, the density of markers on the two maps will differ for some regions. The microsatellite map is more appropriate than the gene-based map for revealing variation in recombination rate using Lyon's method since it consists of many randomly chosen markers mapped in a single intraspecific intercross using both males and females. Moreover, microsatellites appear to be widely and evenly distributed across mammalian genomes (Weber and May 1989, and references therein).

The distribution of interval sizes for the mouse microsatellite map has been described by Dietrich et al. (1994). We take a different approach and attempt to estimate the local rate of

recombination based on the location of markers along the genetic map under the assumption that markers are uniformly distributed on the physical map. Our approach is simple and broadly applicable in situations where a genetic map is available and the assumption of uniformity on the physical map can be justified.

Methods

D. melanogaster. The genetic map of *D. melanogaster* was taken from Lindsley and Zimm (1992). Loci assigned to genetic map positions based solely on their cytogenetic position were excluded. The *D. melanogaster* genome consists of a pair of sex chromosomes and three pairs of autosomes (chromosomes 2, 3 and 4). The fourth chromosome is very small and does not recombine and was excluded from the analysis. There is no recombination in male *Drosophila* so genetic distances reflect female meioses only. The physical lengths of each chromosome (Sorsa 1988) were compared with the genetic lengths of each chromosome, and average recombination rates per chromosome were calculated (in cM / Mb). Observed numbers of markers were compared with the numbers expected based on the physical length of each chromosome. The Kolmogorov-Smirnov statistic was used to test for uniformity in the distribution of markers along the genetic map.

The local density of markers along each chromosome was estimated using a kernel density estimator (Silverman 1986) with a cosine kernel function and 10 cM bandwidth. In order to minimize edge effects we used a reflecting boundary (Silverman 1986, p. 30). For a chromosome of size P in physical units (Mb) and size G in genetic units (cM), we assume that there is a smooth (differentiable) and monotone function F: [0, P] -> [0, G] that maps physical locations along the chromosomes onto genetic locations. The recombination rate (in units of cM / Mb) is defined to be the derivative of this function, d/dt F(t). Now suppose that points x_i are located uniformly at random along the physical map. The corresponding points $y_i = F(x_i)$ on the genetic map will be distributed on the interval [0,

G] with cumulative distribution $F^{-1}(t)$. Thus, the recombination rate is the reciprocal of the density function of these points along the genetic map.

Recombination rates calculated for *D. melanogaster* were compared with the recombination rates reported by Kliman and Hey (1993) and Kindahl and Aquadro (1995) for each chromosome. Kliman and Hey (1993) estimated recombination rates from plots of genetic position versus physical position for markers on each chromosome. They generated a curve for each chromosome by least-squares polynomial curve fitting and estimated recombination rate by taking the derivative of the polynomial. Kindahl and Aquadro (1995) estimated recombination rates by comparing the genetic and cytogenetic distance between many pairs of markers over different genomic regions. Cytogenetic distance was converted to distance in basepairs using the average value of 21 kb per polytene band (Sorsa 1988).

M. musculus. The microsatellite map for *M. musculus* was taken from Release 8 of the Whitehead Institute/MIT genome center genetic map of the mouse (1995). This map includes dense coverage of all 19 autosomes and the X chromosome. As with *Drosophila*, the numbers of markers observed per chromosome were compared with the number expected based on the physical length of each chromosome (Evans 1989). Kolmogorov-Smirnov tests were used to see if the distribution of markers on the microsatellite map differs from a uniform distribution, and density functions were calculated as described above. Because the average density of the mouse map is lower than the average density for the fly map (see Results), we used a 15 cM bandwidth to generate density plots.

We also investigated variation in recombination rate near centromeres and telomeres. Observed numbers of markers within centromeric and telomeric 10 cM regions were compared with expected values under the null hypothesis of random distribution using chi-square tests of homogeneity. Additional inferences about recombination rates near telomeres in *M. musculus* were made from comparisons of the cytogenetic and genetic maps. For each chromosome, a marker in the distal third was chosen which had been

mapped genetically and cytogenetically. The proportion of the cytogenetic map and the proportion of the genetic map distal to that marker were calculated. Cytogenetic distances were calculated from measurements of the standard G-banded idiogram (Evans 1989). Distances distal to the marker included the entire width of the G-band to which the marker was mapped, and thus are overestimates of the true physical distance. Genetic distances distal to the marker include all markers to the end of the genetic map of each chromosome. These are underestimates of the true genetic distance, since they depend on the location of the most distal marker discovered. Comparison of these two proportions provides an underestimate of the relative recombination rate at the telomere compared to the average rate for the chromosome.

Results

D. melanogaster: A total of 2,068 markers span 294 cM, giving an average density of 7.03 markers per cM. Average recombination rates are similar for the second and third chromosomes and are slightly lower than the average for the X chromosome (Table 1). The numbers of markers observed on the three major chromosomes (X, 2, and 3) differ significantly from the numbers expected based on the cytogenetic length of each chromosome ($P < 0.001$, Table 1). This is due to an over-representation of markers on the X chromosome relative to the autosomes, and can be accounted for by the greater attention that researchers have paid to the X because of its simplified genetics. The distributions of markers on the linkage map for each of the three *D. melanogaster* chromosomes differ from random expectations (Figure 1), indicating significant clustering of loci on the genetic map.

The number of markers, density function, and estimated recombination rates for each *D. melanogaster* chromosome are shown in Figure 2. Each chromosome contains one major region in which loci are clustered. For the metacentric chromosomes 2 and 3, these clusters are found at the centromeres, and for the acrocentric X chromosome, the cluster is found at the telomere. Comparison of these major clusters with estimates of recombination

rates for these same regions from Kliman and Hey (1993) and Kindahl and Aquadro (1995) shows that major clusters correspond to regions of lowest recombination, as proposed by Lyon (1976). In addition to the concordance in regions of severely reduced recombination, the three methods are concordant in identifying some but not all of the smaller variations in recombination rate. Discrepancies among the three approaches are most evident at the ends of the chromosomes, where experimental biases and biases due to the density estimation technique are expected.

Scatterplots comparing recombination rates estimated using the different approaches are shown in Figure 3. In general, the best concordance among the three methods is seen for the third chromosome, and the poorest concordance is seen for the X chromosome.

M. musculus: A total of 5,731 markers on the 19 autosomes and the X chromosome cover 1,396 cM, giving an average density of 4.11 markers per cM. Average recombination rates for the different mouse chromosomes are similar, with chromosome 18 showing the lowest rate (0.35 cm / Mb) and chromosome 19 showing the highest rate (0.72 cm / Mb) (Table 2). The numbers of markers observed per chromosome for the microsatellite map differ significantly from the values expected based on the physical length of each chromosome, but do not differ significantly if chromosome 19 and the X chromosome are excluded, as previously noted (Dietrich et al. 1992, 1994). It remains unclear why the X chromosome has about half the expected number of microsatellites, although it may reflect either (a) a lower density of $(CA)_n$ repeats on the X or (b) a lower level of polymorphism on the X, since only polymorphic markers are mapped (Dietrich et al. 1992, 1994). The distribution of markers on each of the 20 *M. musculus* chromosomes differ from random expectations (Figure 4), indicating significant heterogeneity in the distribution of markers along the linkage map. This clustering is also observed in an analysis of the order of crossovers and markers in the microsatellite mapping data (Dietrich et al. 1994).

The number of markers and density function for each *M. musculus* chromosome are shown in Figure 5. The 20 acrocentric mouse chromosomes are shown with the centromeres on the left and the telomeres on the right. Heterogeneity in recombination rate can be seen for most mouse chromosomes, although clear patterns, such as seen for the *D. melanogaster* second and third chromosomes, are not evident in the mouse. In general, clusters of markers tend to fall somewhere in the middle of chromosomes, rather than at the ends, consistent with Lyon's earlier findings.

Telomeric regions in twelve of the 20 chromosomes contain significantly fewer markers than expected under a random distribution (Table 3). Of the remaining eight chromosomes, seven show a trend in the same direction (i.e. observed values less than expected values). These data suggest that recombination rates may be elevated at many or most mouse telomeres. To further explore this observation, a comparison of the gene-based map and cytogenetic map near telomeres is shown in Table 4. For most (but not all) chromosomes, the cytogenetic length is smaller than the genetic length near telomeres, consistent with elevated recombination rates in these regions. Chromosomes 10, 14, and 18 (Table 4) do not show this pattern. This may reflect bias in the calculations, which underestimate genetic distance and overestimate physical distance (see Methods) or it may reflect real differences among individual chromosomes. None of these three chromosomes showed significantly different numbers of markers in telomeric regions than expected, and chromosome 18 is the single autosome for which the observed value is higher than the expected value (Table 3).

Centromeric regions do not show a consistent pattern (Table 3), although four chromosomes showed significantly greater numbers of markers than expected, suggesting a reduction in recombination rate in these areas.

Discussion

D. melanogaster: The distribution of markers along the *D. melanogaster* linkage map suggests variation in regional recombination rate that is broadly concordant with other studies (Lindsley and Sandler 1977; Kliman and Hey 1993; Kindahl and Aquadro, 1995). In particular, the regions of lowest recombination show good agreement between the three methods. In addition, some of the smaller patterns of variation identified by Lyon's method are also concordant with the other studies. For example, the plateau and decline in recombination rate on the X chromosome between positions 40 and 58 are identified by all three methods (Figure 2). There are several examples of minor variation identified by Lyon's method and by Kindahl and Aquadro, but not seen by Kliman and Hey. For example, Kliman and Hey's method does not identify a high peak in recombination rate at position 10 on the X, while the other two methods do (Figure 2). There are also small patterns of variation identified by only one of the three methods and not the others. In general, the concordance between estimates from Lyon's method and estimates from the other two methods is similar to the concordance between estimates from the two other methods compared to each other (Figure 3).

Some discrepancies among the methods are observed at the ends of chromosomes, particularly where gene density is low. The method used here may be biased near the chromosome ends. As the kernel of the density estimator moves beyond the end of the chromosome, there are no markers and thus the local density of markers may be underestimated and the estimated recombination rate will be inflated. We have attempted to minimize the bias by using a reflecting boundary; however, estimates within one bandwidth of the chromosome end should be interpreted with some skepticism. The method of Kindahl and Aquadro may also be less accurate near the chromosome ends. The coefficient of exchange they calculated is an average of many comparisons for each band. Because there are fewer such comparisons for bands at the ends of chromosomes, the standard errors of the estimates are higher in these regions. Both of these factors may help account

for the differences observed at the centromere of the X and the telomeres of chromosomes 2 and 3. In general, the experimental error associated with mapping markers near chromosomes ends (where linkage can be detected in one direction only) may be greater than in other regions.

The smoothness of estimated recombination rates using our implementation of Lyon's method will depend on the choice of bandwidth for the density estimator. Larger bandwidths will produce a smoother picture and smaller bandwidths will reveal more detail of the variation in recombination rate. Significant local variation in recombination rate is likely to exist in most genomes, but the method used here is limited by the total number of markers on the genetic map. If the chosen bandwidth is too small, the density estimates will start to reflect random variation and thus density estimates are less reliable. The estimates presented here should be viewed as average recombination rates over local regions of the chromosome approximately equal to the bandwidth. The bandwidths (10 cM for *Drosophila* and 15 cM for *Mus*) were chosen by a visual assessment over a range of bandwidths. Automated procedures for bandwidth selection are available but there is no general consensus on which is best and the visual method is adequate in most cases (Jones et al. 1995).

The method of Kliman and Hey, in which genetic position is plotted against physical position for each chromosome, is also expected to have a smoothing effect on variation since it is based on fitting a curve to the data. Kindahl and Aquadro's method, which is based on a direct comparison of genetic and physical distance for many pairs of loci, provides the greatest detail, but also depends heavily on the availability of pairs of markers for each genomic region.

One of the early concerns with Lyon's method was that the observed variation in the distribution of markers on a linkage map may reflect either (1) a non-random subset of genetic markers from the genome (e.g. something about how we discover markers leads to clustering), or (2) a non-random physical distribution of markers. If either of these were

very common, then the distribution of markers would not be expected to accurately reflect differences in levels of recombination. For example, a physical cluster of markers would lead to an underestimation of recombination rate using Lyon's approach. Because of the broad concordance seen among the three methods, these alternative hypotheses can be ruled out as general explanations.

M. musculus: Three general observations characterize the mouse density plots. First, most clusters of markers tend to fall somewhere in the middle of the chromosomes, rather than at the ends. This observation is consistent with cytological observations of chiasma formation during meiosis, which often occurs at the ends of the chromosomes (e.g. Polani 1972).

Second, dramatically reduced recombination near centromeres, such as seen for chromosomes 2 and 3 of *D. melanogaster* (Figure 2), is not observed in the mouse data (Figure 5). This could reflect a real difference between *M. musculus* and *D. melanogaster*, but could also be caused by other factors. One possibility is that reduced recombination occurs near mouse centromeres, but that the effect is limited to a small physical region and is therefore undetected. Another possibility is that centromeric regions of mouse chromosomes may contain a lower density of markers per physical distance. While there is no evidence for dramatically reduced recombination at all mouse centromeres, centromeric regions of four chromosomes did contain greater than expected numbers of markers, consistent with lower recombination rates in these intervals (Table 3). In *Drosophila*, strong centromeric suppression of recombination is seen only for the second and third chromosomes, which are metacentric, but not for the X chromosome, which is acrocentric. In the mouse, all 20 chromosomes are acrocentric and none show the degree of suppression seen for the fly metacentric chromosomes. It is possible that, in general, strong centromeric suppression is associated more commonly with metacentric centromeres, than with acrocentric centromeres. Consistent with this hypothesis is the

observation of reduced recombination around the metacentric human X chromosome (Weeks et al. 1995).

Third, recombination appears elevated near telomeres of most mouse chromosomes. This is seen in the distribution of microsatellites (Table 3) and is supported by comparison of the cytogenetic and genetic maps for most chromosomes (Table 4) and by cytological observations of chiasmata in meiosis (Polani 1972). Nonetheless, this trend is not seen for all chromosomes in Tables 3 and 4, and it is possible that real differences exist among chromosomes. Lyon (1976) suggested that for chromosomes which have two chiasmata on average, interference may increase recombination in proximal and distal regions, but for chromosomes with one chiasma on average, recombination may appear more random. Consistent with this hypothesis, several large chromosomes show similar patterns of clustering, with fewer markers in both proximal and distal regions. Elevated rates of recombination near telomeres have previously been suggested for individual mouse chromosomes. For example, Disteche et al. (1989) and Lyon et al. (1987) have compared the physical and genetic maps of the X chromosome and suggested an increase in recombination in the distal portion. Genetic data from both male and female meioses in the F_2 intercross of two inbred mouse strains suggests that recombination rates are elevated near the telomere of chromosome 19 (Dietrich et al. 1994). In humans, elevated rates of recombination near telomeres is suggested by the positions of chiasmata in male meioses for all autosomes (Laurie and Hulten 1985) and has been corroborated for some chromosomes by comparison of physical and genetic maps (e.g. chromosome 21, Tanzi et al. 1992). It is possible that increased recombination near telomeres will turn out to be a general feature of mammalian genomes. This trend is not seen in *D. melanogaster* where the recombination rate is strongly reduced near the telomere of the X chromosome and is only moderately high near the telomeres of chromosomes 2 and 3 (Figure 2).

The recombination rates estimated here for the mouse are based on a (OB x CAST) F_2 intercross. These rates are not necessarily representative of other mouse crosses or of

recombination rates in natural populations of either *Mus domesticus* or *Mus musculus*. In fact, differences in recombination percentages are often observed between the same set of markers among crosses involving different strains of mice (Davisson et al. 1989, and references therein).

The mean recombination rate in the mouse can be estimated by comparing the total genetic and physical length of the genome. The total genetic length is approximately 1,400 cM and the total haploid DNA content is approximately 3×10^9 bp. This corresponds to an average genome-wide recombination rate of 0.47 cM/Mb. Within individual chromosomes, recombination rates vary between fourfold and tenfold. Regions of low recombination are typically in the neighborhood of 0.25 cM / Mb while regions of high recombination can reach 2 cM / Mb or more. Direct measurements of the recombination rates for different regions will be possible over the next two decades as physical maps of the entire genome are assembled as part of the mouse genome project.

Why might recombination rates vary in different genomic regions? On a fine scale, recombination may be facilitated by the presence of particular sequences, including dinucleotide repeats (e.g. Slightom et al. 1980). However, it is unlikely that this fine-scale variation would be detectable over 15 cM regions. On a larger genomic scale, recombination may be affected by mechanical processes. For example, suppression of recombination around the centromeres in Robertsonian translocation heterozygotes (e.g. Cattanach 1978) may extend over large genomic regions (10-40 cM), and appears to be due to mechanical interference resulting in delayed pairing (Davisson and Akeson 1993). Whether or not broad patterns of recombination in genomes without structural heterozygosity will depend on such things as the timing of pairing remains to be seen. Chandley (1986) has proposed that effective pairing (as a prerequisite to recombination) depends on the presence of (and is initiated by) early replicating sequences such as those seen in telomeric regions of mammalian genomes.

There has been much interest recently in relating rates of recombination to levels of genetic variation found in natural populations (e.g. Aguade et al. 1989; Begun and Aquadro 1991; Berry et al. 1991; Begun and Aquadro 1992; Nachman and Aquadro 1993). Theoretical models predict that levels of nucleotide diversity will be lower in regions of little or no recombination, either due to the fixation of adaptive substitutions (Maynard Smith and Haigh 1974; Kaplan et al. 1989) or due to frequent selection against deleterious alleles (Charlesworth et al. 1993). While these predictions have been largely upheld in *D. melanogaster*, testing them in other species, such as mice, has been hampered by our inability to document variation in regional recombination rates.

Gene density and recombination rates: The genome projects for humans, mice, and a few other organisms will provide us with detailed genetic and physical maps over the next two decades, allowing us to measure variation in recombination rates in different genomic regions directly. For most organisms, however, physical maps will probably never exist. The chief utility of Lyon's method is that it provides a means for assessing patterns of recombination when only a genetic map is available. Recently, there has been a rapid proliferation of detailed genetic maps for new species (e.g. Tanksley et al. 1992), owing largely to the ease of isolating highly variable molecular markers such as microsatellites.

When using marker density to evaluate patterns of recombination, several caveats should be kept in mind. First, genetic maps based on randomly chosen markers are preferable to maps based on genes studied for a particular biological reason. It is likely that the latter will contain clusters of markers around particular genes of interest. For example, chromosome 17 in the mouse contains over twice the expected number of markers on the gene-based map and contains the well studied *major histocompatibility complex* and the *t-complex*. It is likely, however, that most new genetic maps will be based on randomly chosen molecular markers. Second, the choice of bandwidth may have an effect on the observed patterns of recombination. In choosing, the appropriate bandwidth for a

particular map, both the average density and the degree of clustering should be considered. Broader bandwidths will produce a more reliable but less detailed picture of heterogeneity in recombination rates. Narrow bandwidths are subject to random fluctuations and may exaggerate heterogeneity. Visual comparison of estimates across a range of bandwidths is recommended.

Acknowledgments

We thank C.F. Aquadro, D.J. Begun, M.J. Ford, M.T. Hamblin, and M.F. Lyon for comments and discussion. This work was supported by the N.S.F. (DEB 9306493).

Literature Cited

Aguade, M., N. Miyashita and C.H. Langley, 1989 Reduced variation in the yellow-achaete-scute region in natural populations of *Drosophila melanogaster*. Genetics 122: 607-615.

Avner, P., L. Amar, L. Dandolo and J.L. Guenet, 1988 Genetic analysis of the mouse using interspecific crosses. Trends in Genet. 4: 18-23.

Begun, D.J., and C.F. Aquadro, 1991 Molecular population genetics of the distal portion of the X chromosome in *Drosophila*: Evidence for genetic hitchhiking of the yellow-achaete region. Genetics 129: 1147-1158.

Begun, D.J., and C.F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356: 519-520.

Berry, A.J., J.W. Ajioka and M. Kreitman, 1991 Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. Genetics: 129: 1111-1117.

Bryda, E.C., J.A. DePari, D.B. Sant'Angelo, D.B. Murphy and H.C. Passmore, 1992 Multiple sites of crossing over within the *Eb* recombinational hotspot in the mouse. Mamm. Genome 2: 123-129.

Cattanach, B.M., 1978 Crossover suppression in mice heterozygous for tobacco mouse metacentrics. Cytogenet. Cell Genet. 20: 264-281.

Chandley, A.C., 1986 A model for effective pairing and recombination at meiosis based on early replicating sites (R-bands) along chromosomes. Hum. Genet. 72: 50-57.

Charlesworth, B., M.T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289-1303.

Copeland, N.G., and N.A. Jenkins, 1991 Development and applications of a molecular genetic linkage map of the mouse genome.

Copeland, N.G., N.A. Jenkins, D.J. Gilbert, J.T. Eppig, L.J. Maltais, J.C. Miller, W.F. Dietrich, A. Weaver, S.E. Lincoln, R.G. Steen, L.D. Stein, J.H. Nadeau and E.S.

Lander, 1993 A genetic linkage map of the mouse: current applications and future prospects. Science 262: 57-66.

Davisson, M.T., and E.C. Akeson, 1993 Recombination suppression by heterozygous Robertsonian chromosomes in the mouse. Genetics 133: 649-667.

Davisson, M.T., T.H. Roderick and D.P. Doolittle, 1989 Recombination percentages and chromosomal assignments, pp. 432-505 in *Genetic variants and strains of the laboratory mouse, second edition*, edited by M.F. Lyon and A.G. Searle. Oxford University Press, Oxford.

Dietrich, W.F., H. Katz, S.E. Lincoln, H-S Shin, J. Friedman, N.C. Dracopoli and E.S. Lander, 1992 A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131: 423-447.

Dietrich, W.F., J.C. Miller, R.G. Steen, M. Merchant, D. Damron, R. Nahf, A. Gross, D.C. Joyce, M. Wessel, R.D. Dredge, A. Marquis, L.D. Stein, N. Goodman, D.C. Page and E.S. Lander, 1994 A genetic map of the mouse with 4,006 simple sequence length polymorphisms. Nature Genetics 7: 220-225.

Disteche, C.M., G.K. McConnell, S.G. Grant, D.A. Stephenson, V.M. Chapman, S. Gandy and D.A. Adler, 1989 Comparison of the physical and recombination maps of the mouse X chromosome. Genomics 5: 177-184.

Evans, E.P., 1989 Standard idiogram, pp. 576-578 in *Genetic variants and strains of the laboratory mouse, second edition*, edited by M.F. Lyon and A.G. Searle. Oxford University Press, Oxford.

Green, M.C., 1975. Communication. Mouse News Letter 53, 10.

Grimm, T., B. Muller, M. Dreier, E. Kind, T. Bettecken, G. Meng and C.R. Muller, 1989 Hot spot of recombination within *DXS164* in the Duchenne Muscular Dystrophy gene. Am. J. Hum. Genet. 45: 368-372.

Jones, M.C., J.S. Marron and S.J. Sheather, 1995 A brief survey of bandwidth selection for density estimation. Working paper series; Australian graduate school of management: 95-005.

Kaplan, N.L., R.R. Hudson and C.H. Langley, 1989 The "hitchhiking effect" revisited. Genetics 123: 887-899.

Kindahl, E.C., and C.F. Aquadro, 1995. The recombinational landscape of the *Drosophila melanogaster* genome. Genetics, in revision.

Kliman, R.M., and J. Hey, 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. 10: 1239-1258.

Laurie, D.A., and M.A. Hulten, 1985 Further studies on bivalent chiasma frequency in human males with normal karyotypes. Ann. Hum. Genet. 49: 189-201.

Lebo, R.U., A. Chakravarti, K.H. Buetow, M.C. Cheong, H. Cann, B. Cordell and H. Goodman, 1983 Recombination within and between the human insulin gene and β-globin gene loci. Proc. Nat. Acad. Sci. USA 80: 4808-4812.

Lindsley, D.L., and L. Sandler, 1977 The genetic analysis of meiosis in female *Drosophila melanogaster*. Phil. Trans. R. Soc. Lond. B. 277: 295-312.

Lindsley, D.L., and G.G. Zimm, 1992 *The genome of Drosophila melanogaster*. Academic Press, San Diego.

Lyon, M.F., 1976 Distribution of crossing-over in mouse chromosomes. Genet. Res. Camb. 28: 291-299.

Lyon, M.F., J. Zenthon, M.D. Burtenshaw and E.P. Evans, 1987 Localization of the Hprt locus by in-situ hybridization and distribution of loci on the mouse X-chromosome. Cytogenet. Cell Genet. 44: 163-166.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23-35.

Nachman, M.W., and C.F. Aquadro, 1993 Polymorphism and divergence at the 5' flanking region of the sex-determining locus, Sry, in mice. Mol. Biol. Evol. 11: 539-547.

Polani, P.E., 1972 Centromere localization at meiosis and the position of chiasmata in the male and female mouse. Chromosoma 36: 343-374.

Silverman, B.W., 1986 *Density estimation for statistics and data analysis.* Chapman and Hall, London.

Slightom, J.L., A.E. Blechl, and O. Smithies, 1980 Human fetal $^{G}\gamma$- and $^{A}\gamma$- globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 21: 627-638.

Sokal, R.R., and F.J. Rohlf, 1981 *Biometry, second edition.* W.H. Freeman and Co., New York.

Sorsa, V., 1988 *Chromosome maps of Drosophila.* CRC Press, Inc., Boca Raton, Florida

Steinmetz, M., D. Stephan and K.Fischer Lindahl, 1986 Gene organization and recombinational hotspots in the murine major histocompatibility complex. Cell 44: 895-904.

Tanksley, S.D., M.W. Ganal, J.P. Prince, M.C. de Vicente, M.W. Bonierbale, P. Broun, T.M. Fulton, J.J. Giovannoni, S. Grandillo, G.B. Martin, R. Messeguer, J.C. Miller, L. Miller, A.H. Paterson, O. Pineda, M.S. Roder, R.A. Wing, W. Wu and N.D. Young, 1992 High density molecular linkage maps of the tomato and potato genomes. Genetics 132: 1141-1160.

Tanzi, R.E., P.C. Watkins, G.D. Stewart, N.S. Wexler, J.F. Gusella and J.L. Haines, 1992 A genetic linkage map of human chromosome 21: analysis of recombination as a function of sex and age. Am. J. Hum. Genet. 50: 551-558.

Tautz, D., 1989 Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nuc. Acid. Res. 17: 6463-6471.

Valdes, A.M., M. Slatkin and N.B. Freimer, 1993 Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133: 737-749.

Weber, J.L., and P. E. May, 1989 Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44: 388-396.

Weeks, D.E., T.G. Nygaard, M. Neystat, L.D. Harby and K.C. Wilhelmsen, 1995 A High-resolution genetic linkage map of the pericentromeric region of the human X chromosome. Genomics 26: 39-46.

Whitehead Institute/MIT Genome Center Genetic Map of the Mouse, 1995. Release 8.0.

Table 1. Observed and expected numbers of loci on the three major *D. melanogaster* chromosomes.

Chromo-some	Physical length (Mb)	Genetic length (cM)	Average recombination rate (cM / Mb)	No. of markers observed	No. of markers expected[a]
X	21.1	73.1	3.46	869	417.3
2	40.6	110.0	2.71	624	799.5
3	47.1	110.9	2.35	573	849.1

[a] Expected values are based on the cytogenetic lengths of each chromosome (Lindsley and Zimm 1992). Observed and expected values are significantly different ($\chi^2 = 233.9$, df = 2, $P < 0.0001$).

Table 2. Observed and expected numbers of loci on the 20 *M. musculus* chromosomes.

Chromo-some	Physical length (Mb)	Genetic length (cM)	Average recombination rate (cM / Mb)	No. of markers observed	No. of markers expected[a]
1	216.0	115.6	0.54	458	412.6
2	208.5	97.3	0.47	452	398.3
3	179.7	68.5	0.38	312	343.3
4	176.7	73.4	0.42	305	337.6
5	170.4	84.6	0.50	370	325.5
6	165.9	64.1	0.39	332	316.9
7	155.7	69.6	0.45	322	297.4
8	149.1	70.6	0.47	315	284.8
9	143.7	69.1	0.48	299	274.5
10	142.2	74.2	0.52	267	271.6
11	141.6	83.2	0.59	327	270.5
12	146.4	59.8	0.41	250	279.7
13	131.4	59.2	0.45	282	251.0
14	133.8	64.3	0.48	238	255.6
15	121.5	63.2	0.52	245	232.1
16	114.3	55.0	0.48	198	218.4
17	115.8	51.9	0.45	240	221.2
18	116.4	40.4	0.35	213	222.4
19	81.9	59.1	0.72	111	156.5
X	186.9	72.8	0.39	195	357.0

[a] Expected values are based on the cytogenetic lengths of each chromosome (Evans 1989). Observed and expected values are significantly different ($\chi^2 = 82.73$, df = 19, P = 0.0001).

Table 3. Observed[a] and expected[b] numbers of markers in centromeric and telomeric regions.

	Centromeric Regions			Telomeric Regions		
Chromosome	Observed	Expected	χ^2	Observed	Expected	χ^2
1	33	39.6	1.20	21	39.6	9.59
2	46	46.5	0.01	16	46.5	22.26
3	31	45.5	5.41	30	45.5	6.20
4	40	41.6	0.07	22	41.6	10.66
5	44	43.7	0.00	33	43.7	2.99
6	85	51.8	25.21	15	51.8	30.97
7	40	46.2	0.97	32	46.2	5.12
8	52	44.6	1.43	25	44.6	10.06
9	46	43.3	0.20	15	43.3	21.62
10	41	36.0	0.80	26	36.0	3.21
11	40	39.3	0.01	31	39.3	2.00
12	53	41.8	3.60	21	41.8	12.44
13	68	47.7	10.40	13	47.7	30.35
14	52	37.0	7.20	29	37.0	2.05
15	22	38.8	8.64	31	38.8	1.85
16	34	36.0	0.14	25	36.0	4.09
17	76	46.3	23.61	36	46.3	2.82
18	46	52.8	1.16	56	52.8	0.26
19	17	18.8	0.21	6	18.8	10.47
X	24	26.8	0.34	20	26.8	1.99

[a]Observed values are counted directly from the data and do not rely on the kernel density estimator; these values are thus free of the potential bias at chromosomes ends in Figure 5.
[b]Expected values are based on the null hypothesis of uniform distribution.

Table 4. Comparison of genetic and cytogenetic map lengths in telomeric regions of mouse chromosomes.

Chromosome	Marker	Position (cM)	Proportion of map distal to marker (%)	
			Genetic	Cytogenetic
1	T190Ca	69	22	7
2	T2Wa	76	25	10
3	T24H	83	17	13
4	T1Sn	64	15	14
5	*Tcf1*	57	39	30
6	T32H	70	17	12
7	T7Ad	51	43	16
8	T17H	53	35	15
9	*Cck*	68	18	4
10	T12RI	65	17	25
11	*Apoh*	58	26	17
12	*Aat*	49	35	20
13	T70H	45	38	22
14	T6Ca	44	8	14
15	*Acr*	46	25	9
16	None available			
17	T138Ca	22	58	42
18	T18H	39	35	50
19	T145H	33	33	23
X	*DXWas31*	68	27	16

For each chromosome, a marker was chosen for which reliable cytogenetic and genetic positions are established. The genetic distances distal to the chosen marker are underestimates since their length depends on the number of markers found more distally. The cytogenetic distances distal to the chosen marker are overestimates since they include the entire G-band in which the chosen marker is assigned. Cytogenetic distances are taken from measurements of the standard idiogram of Evans (1989).

Figure Legends

Figure 1. Cumulative distribution of markers along the *D. melanogaster* X, second, and third chromosomes. The distance between successive points on the x-axis is 1/N, where N is the total number of markers for the chromosome. The y-axis depicts the normalized genetic position for each marker. The identity line (x = y) gives the expectation under a uniform physical distribution and no heterogeneity in recombination rate and is shown for reference. The Kolmogorov-Smirnov statistic measures the largest deviation of the cumulative distribution from this line. For each chromosome, this null hypothesis (x = y) is rejected ($P < 0.01$ for each). Under the assumption of uniform physical distribution, the derivative of the curve provides an estimate of recombination rate.

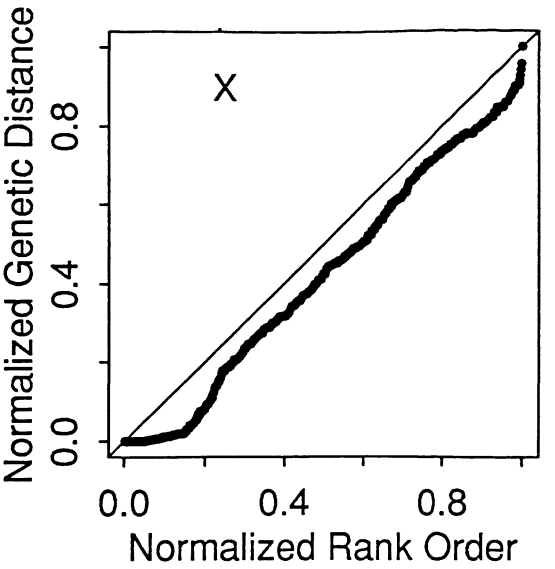
Figure 2. Distribution of markers, estimated density function, and estimated recombination rate for the *D. melanogaster* X (a), second (b) , and third (c) chromosomes. Triangles indicate recombination rates from Kindahl and Aquadro (1995) and circles indicate recombination rates from Kliman and Hey (1993); these are average values per 2.5 cM interval of the map. Horizontal bars indicate the regions of the estimated density that are within one bandwidth of the edge of the chromosome and which may be biased.

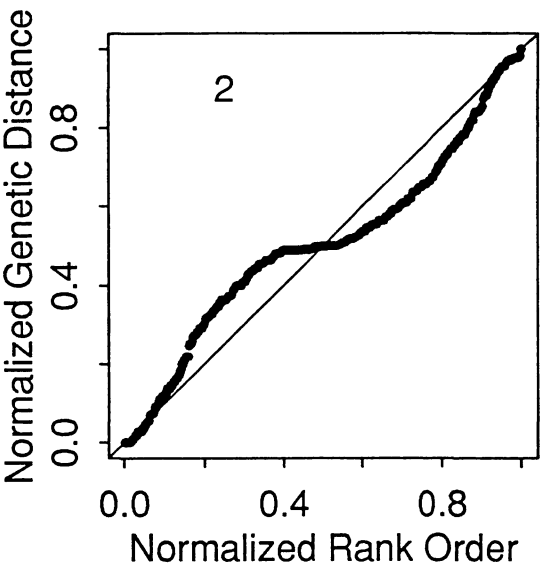
Figure 3. Scatterplots of average recombination rate (cm / Mb) per 2.5 cM interval estimated by three different studies for each of the three major *D. melanogaster* chromosomes. The identity line (x = y) is shown for reference. Numbers shown are correlation coefficients. Estimated densities within one bandwidth of chromosome edges, which may be biased, are excluded.

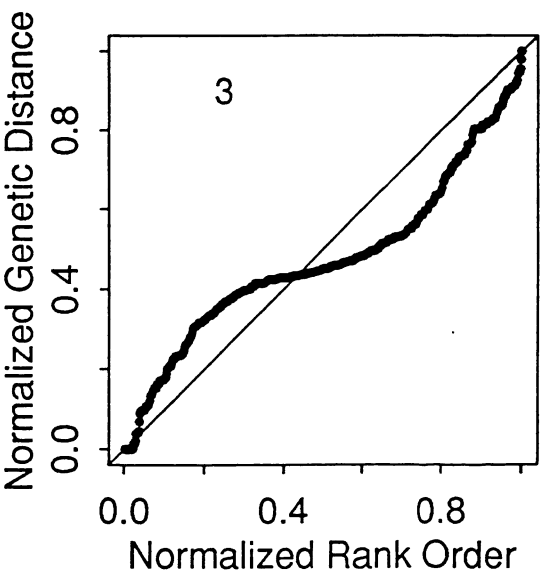
Figure 4. Cumulative distribution of markers along each of the *M. musculus* chromosomes. X-axis: normalized rank order; Y-axis: normalized genetic position. See

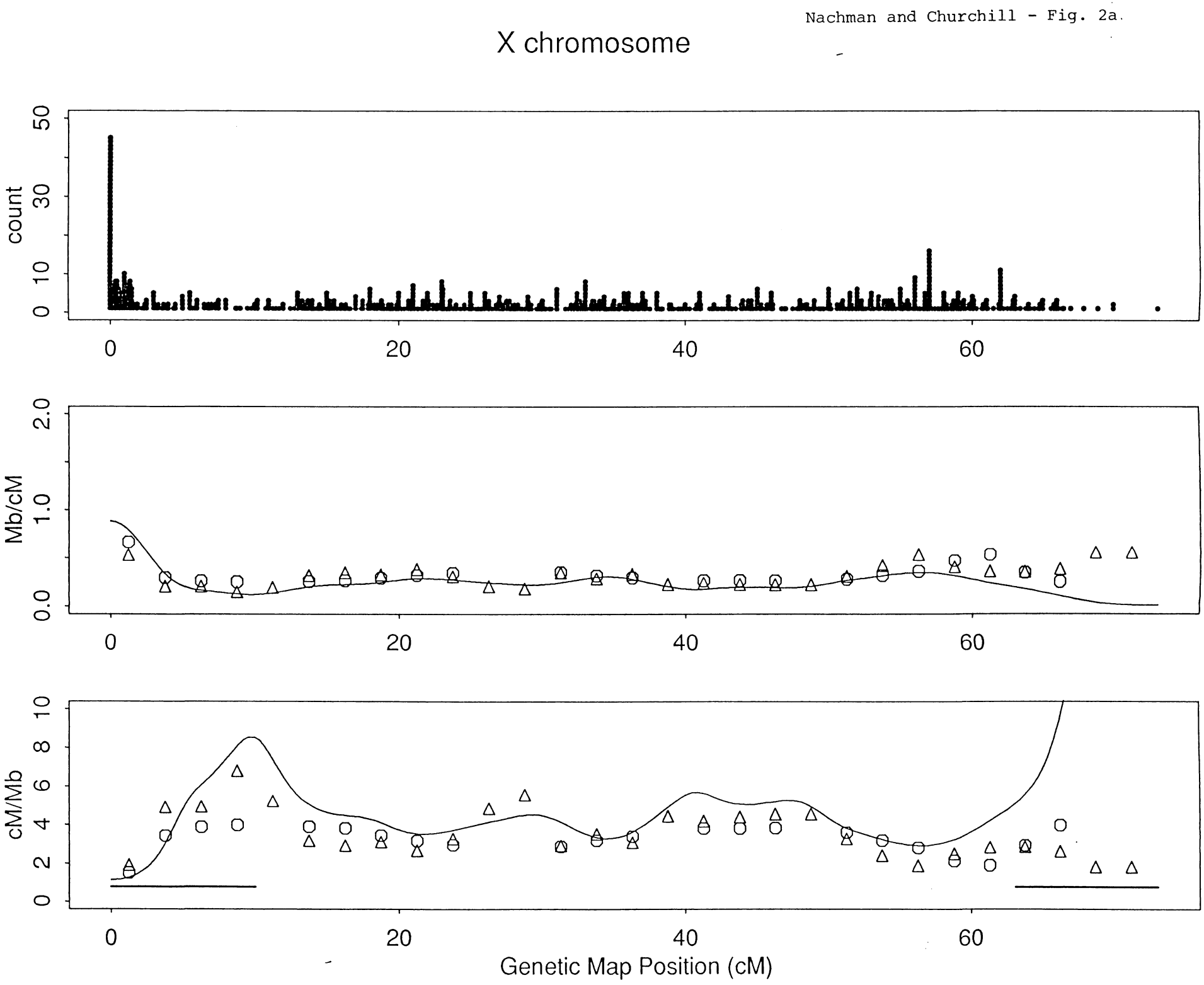
legend to Figure 1. Kolmogorov-Smirnov tests for each chromosome reject the null hypothesis of uniform distribution ($P < 0.01$ for each).

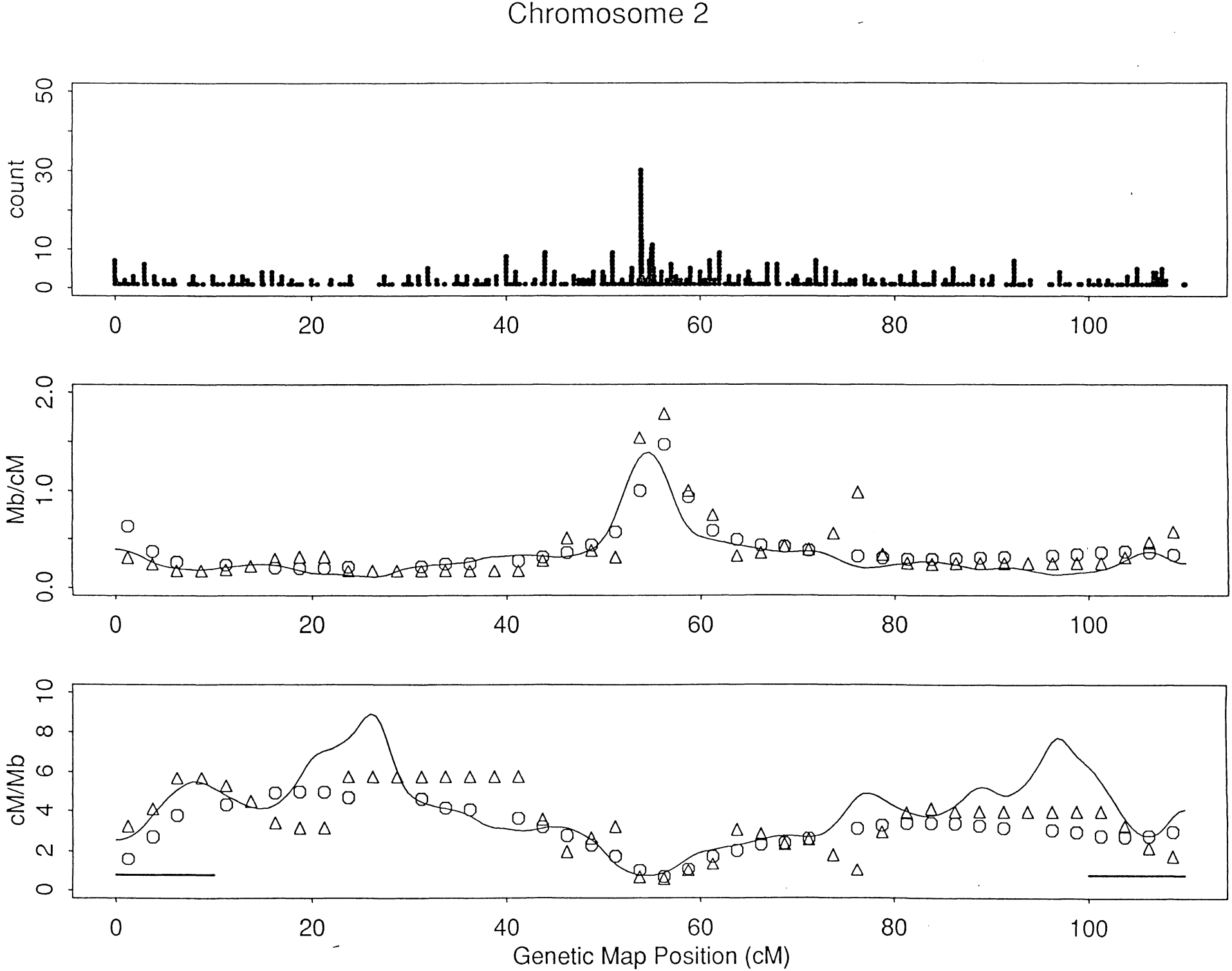
Figure 5. Distribution of markers and estimated density functions for each of the 20 *M. musculus* chromosomes. X-axis is the genetic map position (cM). All chromosomes are acrocentric with centromeres at the left and telomeres at the right.

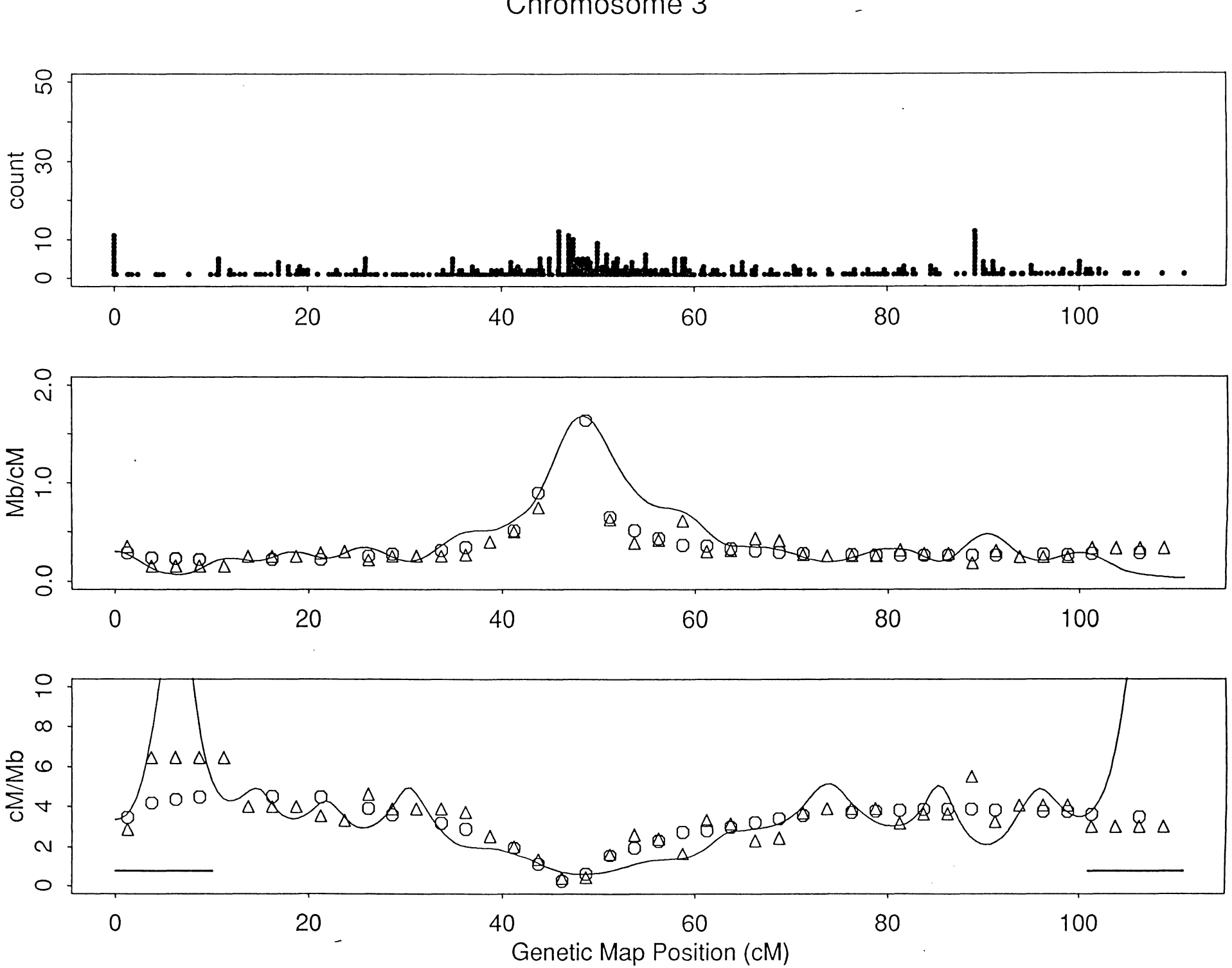

X
Normalized Genetic Distance
0.0
0.4
0.8
0.0
0.4
0.8
Normalized Rank Order


2
Normalized Genetic Distance
0.0
0.4
0.8
0.0
0.4
0.8
Normalized Rank Order

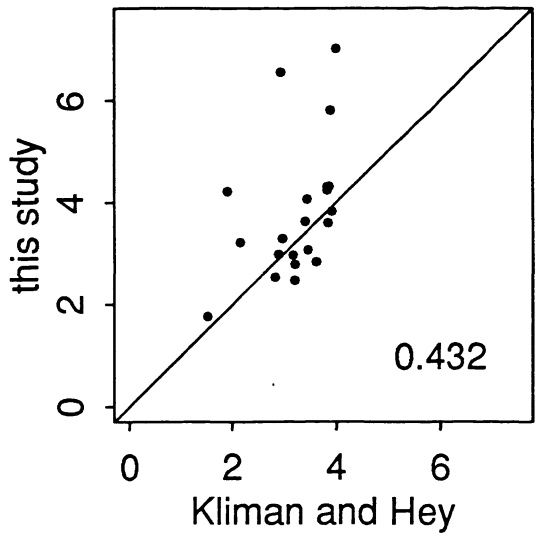

3
Normalized Genetic Distance
0.0
0.4
0.8
0.0
0.4
0.8
Normalized Rank Order

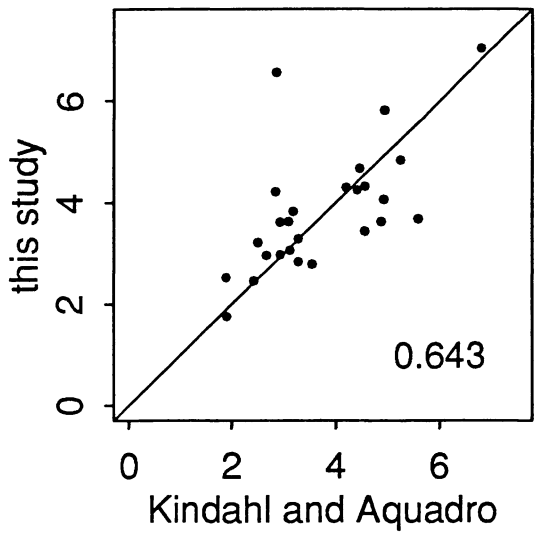

X chromosome
count
0
10
30
50
0
20
40
60
Mb/cM
0.0
1.0
2.0
0
20
40
60
cM/Mb
0
2
4
6
8
10
0
20
40
60
Genetic Map Position (cM)

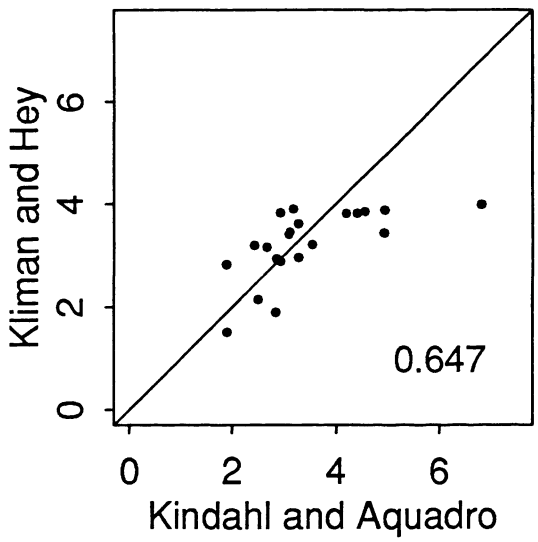

Chromosome 2
count
0
10
30
50
0
20
40
60
80
100
Mb/cM
0.0
1.0
2.0
cM/Mb
0
2
4
6
8
10
Genetic Map Position (cM)


Chromosome 3
count
0
10
30
50
0
20
40
60
80
100
Mb/cM
0.0
1.0
2.0
0
20
40
60
80
100
cM/Mb
0
2
4
6
8
10
0
20
40
60
80
100
Genetic Map Position (cM)

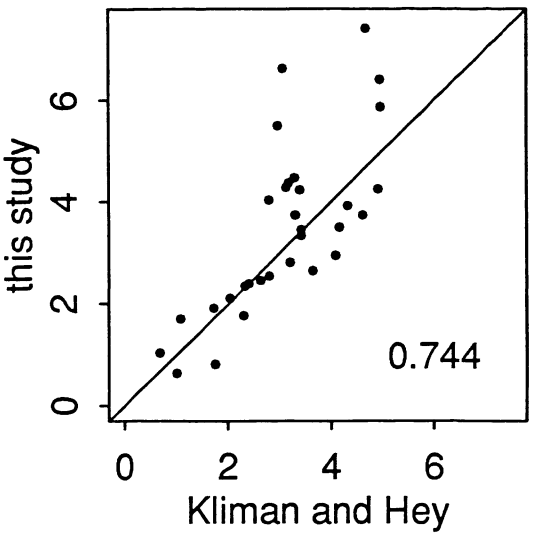
X Chromosome

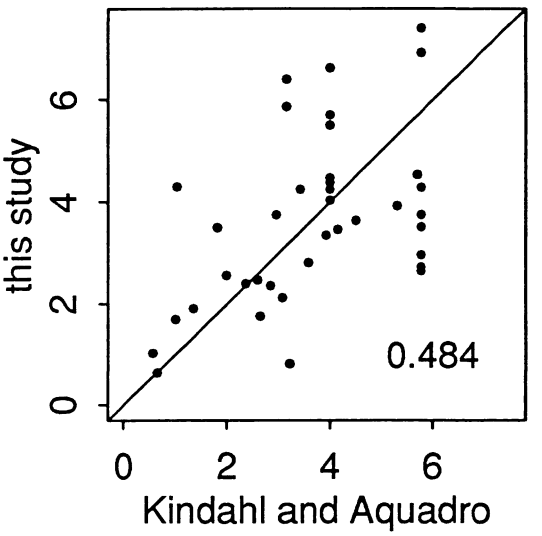


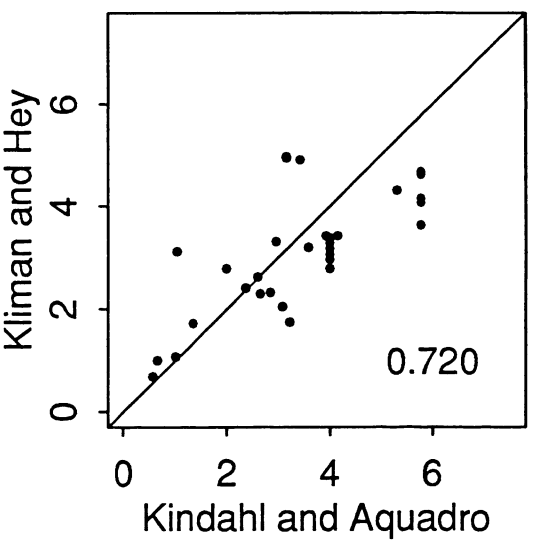




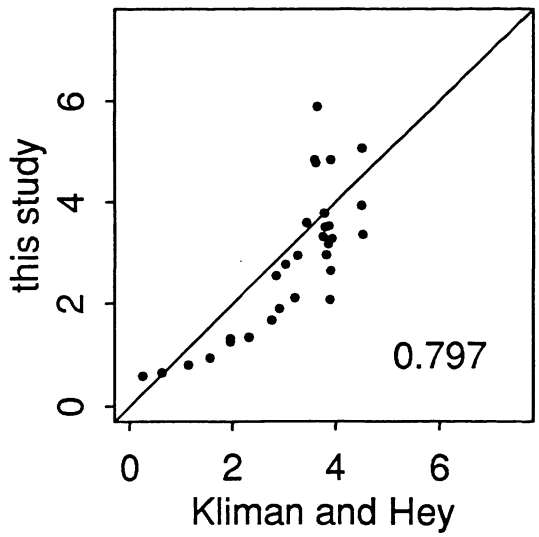
Chromosome 2

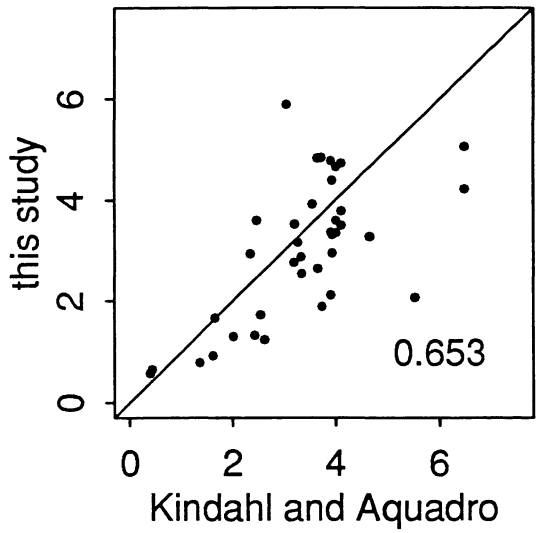


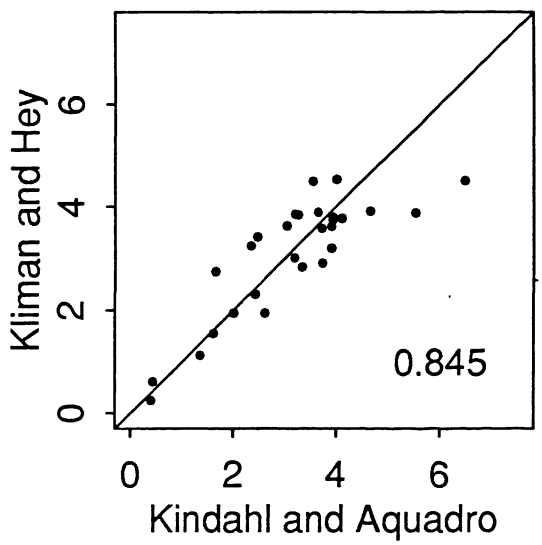




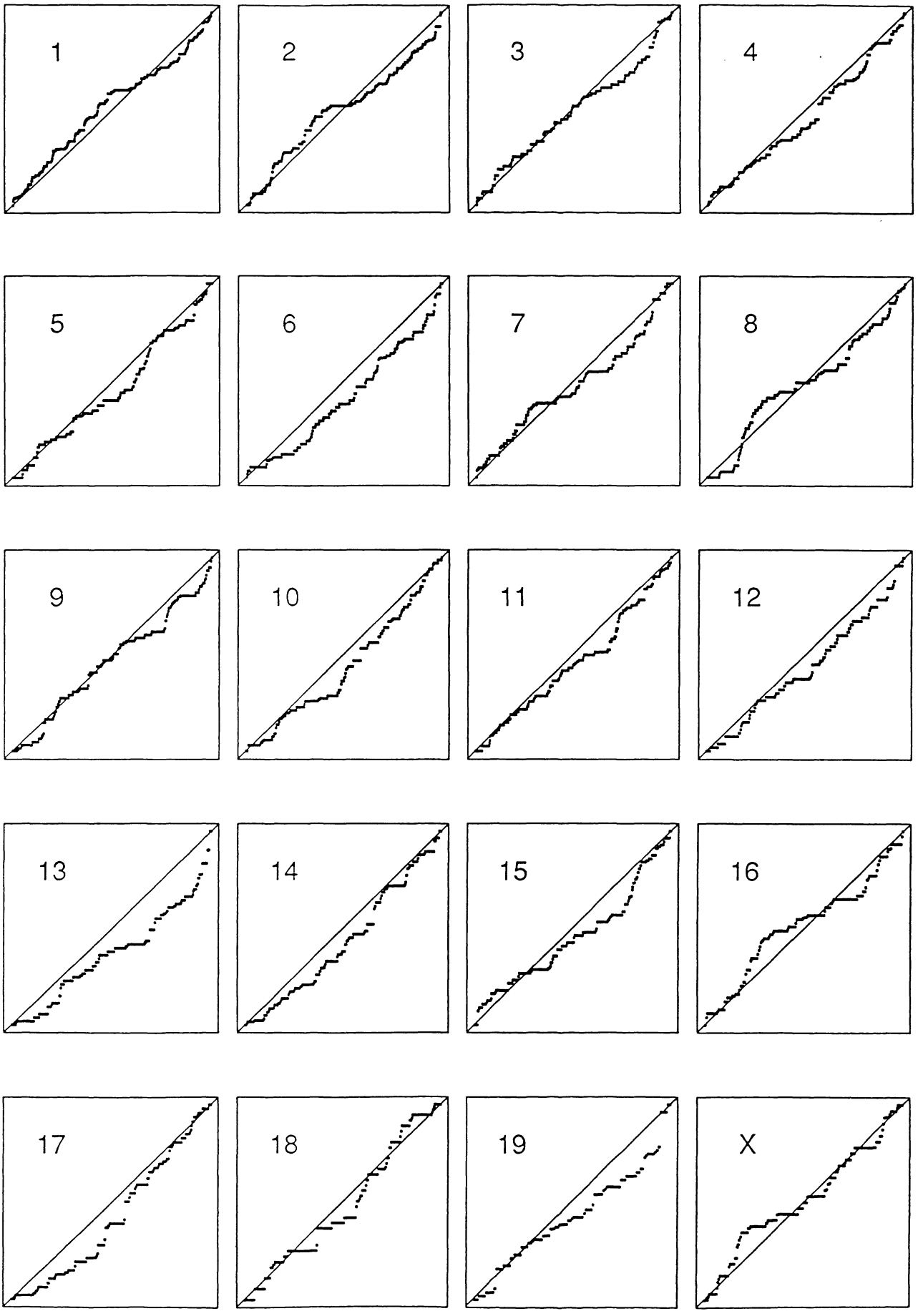
Chromosome 3

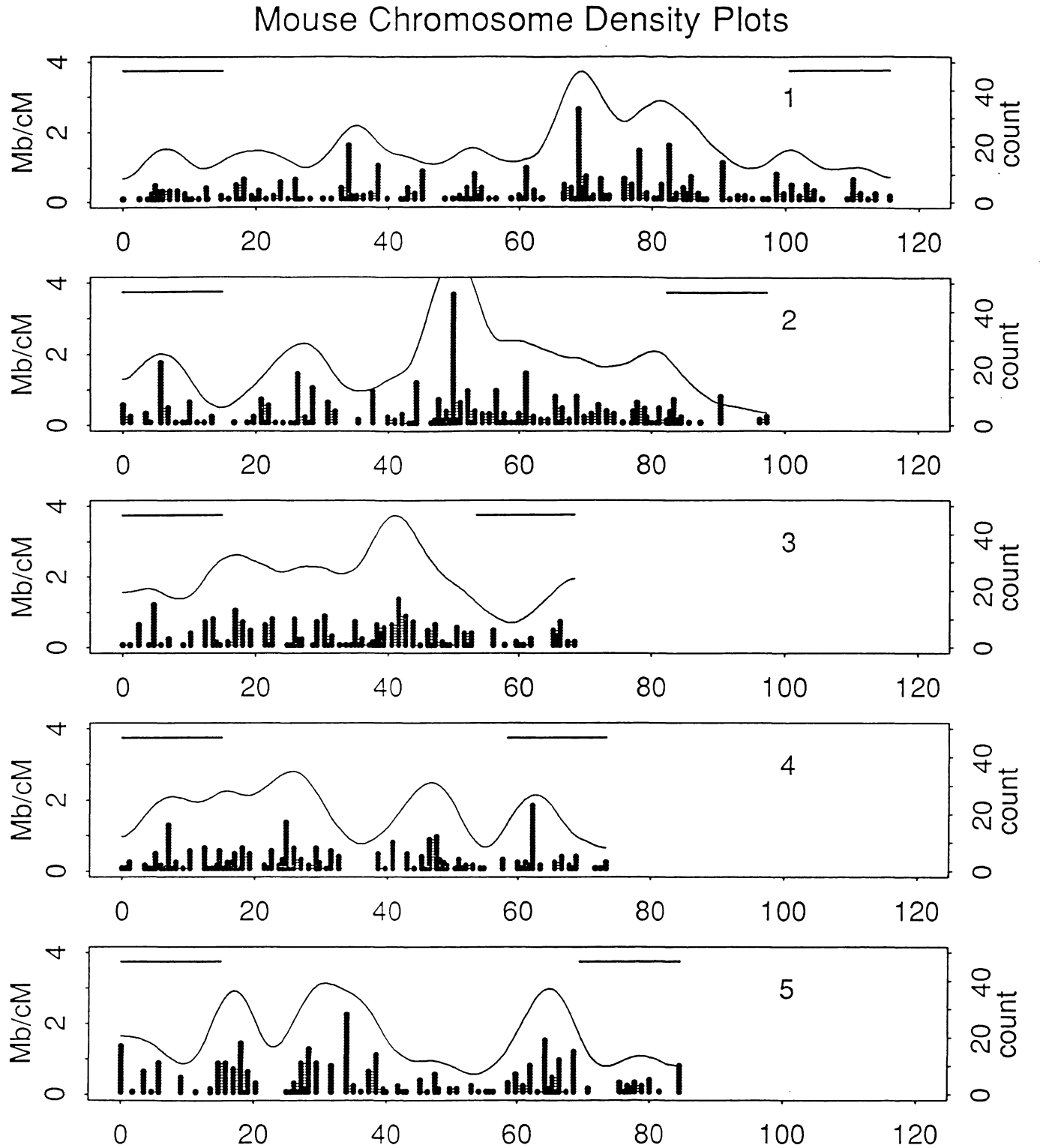


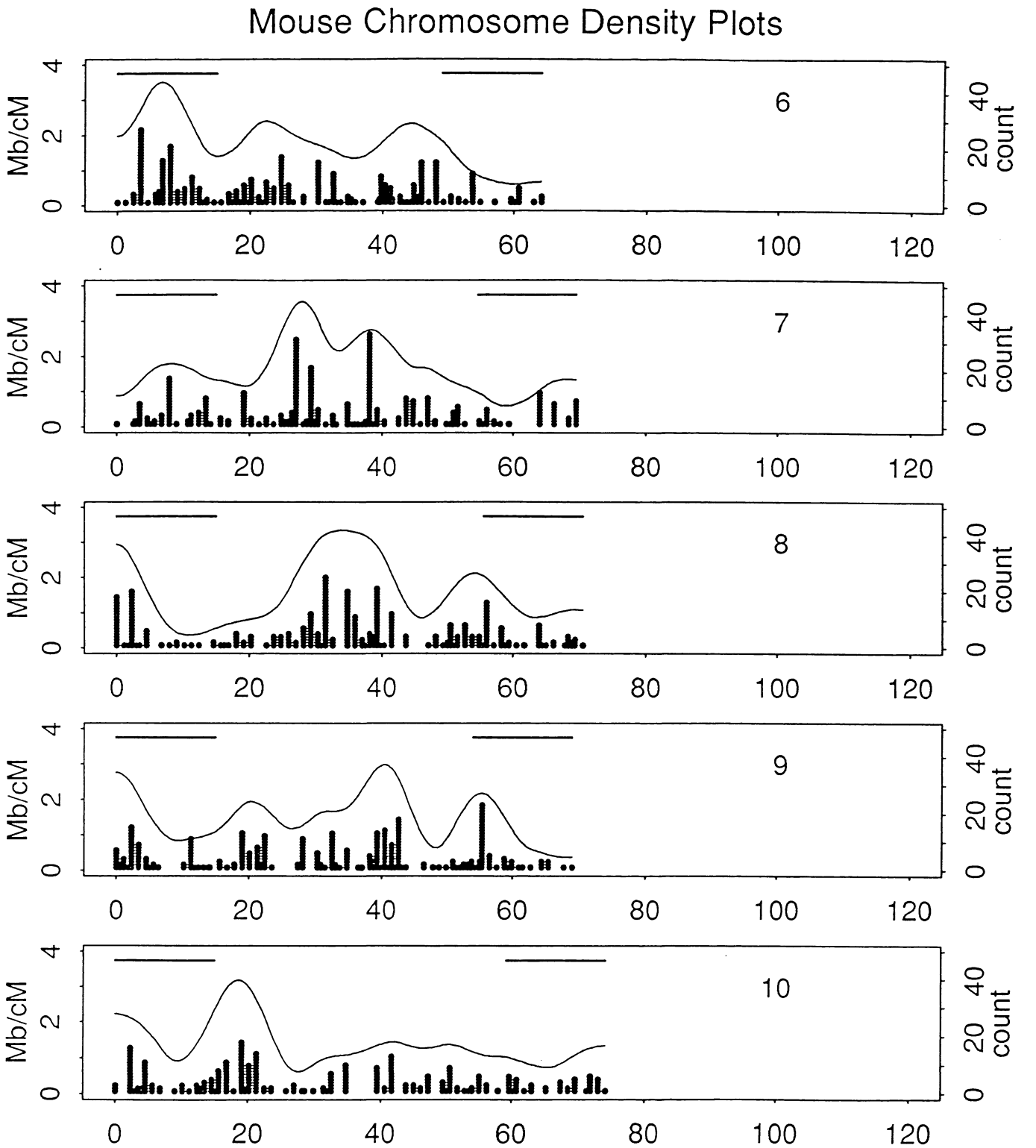


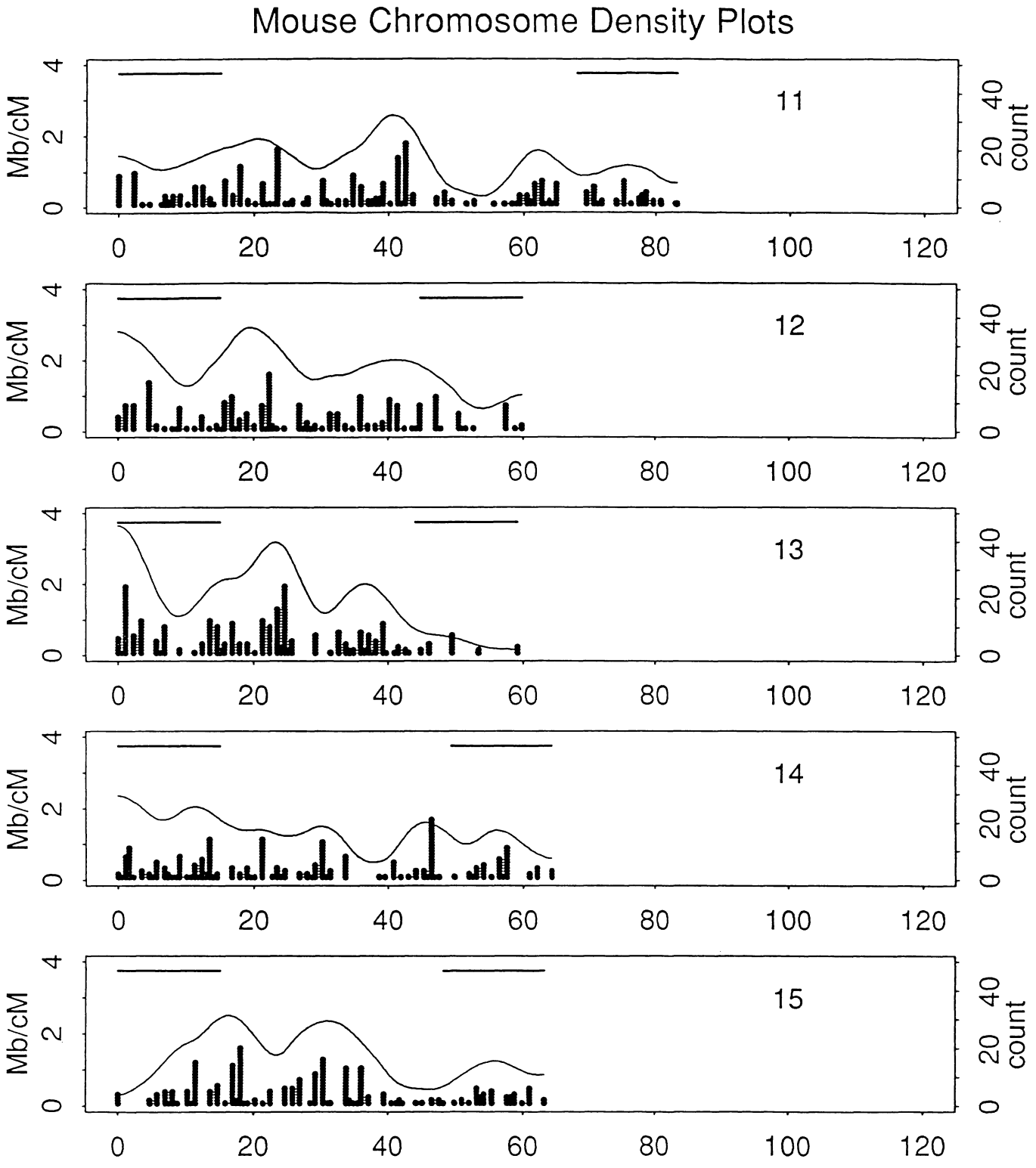


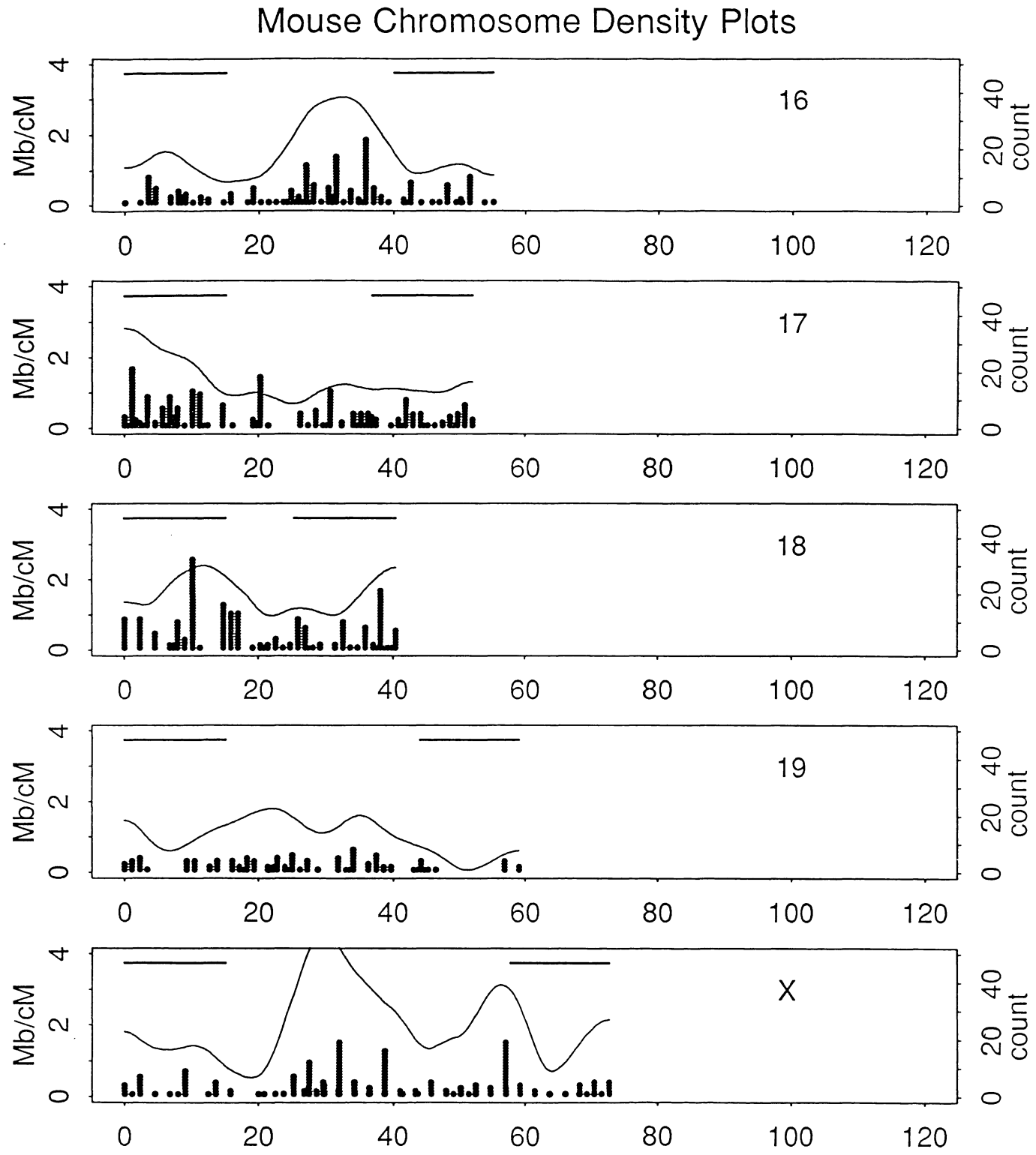
Mouse Chromosome K-S Plots




Mouse Chromosome Density Plots
Mb/cM
count
1
2
3
4
5
0
20
40
60
80
100
120


Mouse Chromosome Density Plots
Mb/cM
count
6
7
8
9
10
0
2
4
20
40
60
80
100
120


Mouse Chromosome Density Plots
Mb/cM
count
11
12
13
14
15
0
2
4
20
40
60
80
100
120


Mouse Chromosome Density Plots
Mb/cM
count
16
17
18
19
X
0
2
4
20
40
60
80
100
120